6. The F Distribution

In this section we will study a distribution that has special importance in statistics. In particular, this distribution arises form ratios of sums of squares when sampling from a normal distribution.

The Density Function

Suppose that U has the chi-square distribution with m degrees of freedom, V has the chi-square distribution with n degrees of freedom, and that U and V are independent. Let

$$X = \frac{U/m}{V/n}$$

1. Show that X has the probability density function

$$f(x) = \frac{\Gamma((m+n)/2)}{\Gamma(m/2)\,\Gamma(n/2)} \left(\frac{m}{n}\right)^{m/2} \frac{x^{(m-2)/2}}{(1 + (m/n)\,x)^{(m+n)/2}}, \quad x \ge 0$$

The distribution defined by the density function in Exercise 1 is known as the ***F* distribution** with m degrees of freedom in the numerator and n degrees of freedom in the denominator. The F distribution was first derived by George Snedecor, and is named in honor of Sir Ronald Fisher.

2. In the random variable experiment, select the F distribution. Vary the parameters with the scroll bars and note the shape of the density function. For selected values of the parameters, run the simulation 1000 times, updating every 10 runs, and note the apparent convergence of the empirical density function to the true density function.

3. Sketch the graph of the F density function given in Exercise 1. In particular, show that

a. f at first increases and then decreases, reaching a maximum at the mode $x = \frac{m-2}{m\,(n+2)}$
b. $f(x) \to 0$ as $x \to \infty$

Thus, the F distribution is unimodal but skewed.

The distribution function and the quantile function do not have simple, closed-form representations. Approximate values of these functions can be obtained from the quantile applet and from most mathematical and statistical software packages.

4. In the quantile applet, select the F distribution. Vary the parameters and note the shape of the density function and the distribution function. In each of the following cases, find the median, the first and third quartiles, and the interquartile range.

a. $m = 5$, $n = 5$
b. $m = 5$, $n = 10$
c. $m = 10$, $n = 5$
d. $m = 10$, $n = 10$

Moments

Suppose that X has the F distribution with m degrees of freedom in the numerator and n degrees of freedom in the denominator. The random variable representation in Exercise 1 can be used to find the mean, variance, and other moments.

5. Show that

a. $\mathbb{E}(X) = \infty$ if $n \in (0, 2]$
b. $\mathbb{E}(X) = \frac{n}{n-2}$ if $n \in (2, \infty)$

Thus, the mean depends only on the degrees of freedom in the denominator.

6. Show that

a. $\text{var}(X)$ is undefined if $n \in (0, 2]$
b. $\text{var}(X) = \infty$ if $n \in (2, 4]$
c. If $n \in (2, \infty)$ then

$$\text{var}(X) = 2\left(\frac{n}{n-2}\right)^2 \frac{m+n-2}{m\,(n-4)}$$

7. In the simulation of the random variable experiment, select the F distribution. Vary the parameters with the scroll bar and note the size and location of the mean/standard deviation bar. For selected values of the parameters, run the simulation 1000 times, updating every 10 runs, and note the apparent convergence of the empirical moments to the true moments.

8. Show that

a. $\mathbb{E}\left(X^k\right) = \infty$ if $n \in (0, 2\,k]$
b. If $n \in (2\,k, \infty)$ then

$$\mathbb{E}\left(X^k\right) = \frac{\Gamma((m+2\,k)/2)\,\Gamma((n-2\,k)/2)}{\Gamma(m/2)\,\Gamma(n/2)}\left(\frac{n}{m}\right)^k$$

Transformations

9. Suppose that X has the F distribution with m degrees of freedom in the numerator and n degrees of freedom in the denominator. Show that $\frac{1}{X}$ has the F distribution with n degrees of freedom in the numerator and m degrees of freedom in the denominator.

10. Suppose that T has the t distribution with n degrees of freedom. Show that $X = T^2$ has the F distribution with 1 degree of freedom in the numerator and n degrees of freedom in the denominator.